STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)

HT2019

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2019-09-23

Skrivtid:	kl. 15.00 - 20.00
Godkända hjälpmedel:	Miniräknare utan lagrade formler och text
Bifogade hjälpmedel:	Häftet *Formelsamling och Tabeller över statistiska fördelningar* (återlämnas efter skrivningen)

- Tentamen består av 7 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.
- **Uppgift 1 – 5**: Svar lämnas på särskild **SVARSBILAGA**,
  - Totalt 12 flervalsfrågor där ett av fem alternativ är korrekt svar.
  - Har fler än ett svarsalternativ markerats för en deluppgift ges noll poäng.
  - Uträkningar lämnas ej in för dessa, om uträkningar ändå lämnas in kommer de inte att beaktas vid bedömningen.
- **Uppgift 6 – 7**: Svar med **FULLSTÄNDIGA REDOVISNINGAR** ska lämnas in.
  - Använd endast skrivpapper som tillhandahålls i skrivsalen.
  - För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
  - Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan också ge poängavdrag!
- Tentamen kan maximalt ge 60 + 40 = 100 poäng och för godkänt resultat krävs minst 50.
- Betygsgränser:

  A: 90 – 100 p
  B: 80 – 89 p
  C: 70 – 79 p
  D: 60 – 69 p
  E: 50 – 59 p
  Fx: 40 – 49 p
  F: 0 – 40 p

  OBS! Fx och F är underkända betyg som kräver omtentamen. Studenter som får betyget Fx kan alltså inte komplettera för högre betyg.
- Lösningsförslag läggs ut på Mondo kort efter tentamen.

LYCKA TILL!

Endast svar ska anges för uppgifterna 1-5 med deluppgifter. Markera rätt svar på Svarsbilagan. OBS! Om du inte hittar rätt svar bland de givna svarsalternativen eller anser att något är fel med uppgiften, skriv ditt svar på Svarsbilagan vid den aktuella frågan och kommentera på baksidan av bladet.

Uppgift 1

Försäkringskassan undersökte hur nöjda småbarnsföräldrar var med kassans hantering av VAB-ärenden (VAB = vård av barn). Man samlade bl.a. in uppgifter om föräldrarnas ålder (i år) och civilstånd (sex klasser), barnens kön och ålder (i månader) och antal barn i hushållet. Varje respondent fick också ange ett omdöme dvs. ange ett betyg kring hur de upplevde hanteringen på en fem-gradig betygsskala där 1 är sämst och 5 högst betyg.

a) Vilket av följande alternativ är ett felaktigt påstående? (5p)

A. Ålder är här en diskret kvantitativ variabel på kvotskala

B. Civilstånd är här en kvantitativ variabel på intervallskala

C. Antal barn är en diskret kvantitativ variabel på kvotskala

D. Kön är en kvalitativ variabel på nominalskala

E. Betyg är en kvalitativ variabel på ordinalskala

Vid en sammanställning av statistiken från undersökningen över antal barn (per kvinna) fick man följande frekvenser (dvs. antal kvinnor med 0, 1, 2, … barn) för totalt $n = 100$ kvinnor:

Antal barn	0	1	2	3	4	5	6	≥ 7
Frekvens	20	28	27	16	6	3	0	0

b) Ange 1:a kvartilen (Q_1) och 3:e kvartilen (Q_3) över antal barn per kvinna. Använd de definitioner för kvartiler som har behandlats i undervisningen och i kurslitteraturen. TIPS: Ta fram den kumulativa fördelningen över antal barn per kvinna. (5p)

A. $Q_1 = 1{,}00$ $Q_3 = 3{,}00$

B. $Q_1 = 1{,}25$ $Q_3 = 1{,}00$

C. $Q_1 = 1{,}00$ $Q_3 = 2{,}75$

D. $Q_1 = 1{,}25$ $Q_3 = 2{,}25$

E. $Q_1 = 2{,}00$ $Q_3 = 2{,}00$

Uppgift 2

En bilhandlare för ett visst bilmärke samlar in uppgifter om nybilsförsäljningar. Det visade sig att 50 % av kunderna hade ägt eller ägde en bil av samma märke ($Å$ = återkommande kunder), 40 % ägde eller hade ägt bil tidigare men aldrig av det märket (N = nya kunder). De återstående 10 % hade aldrig tidigare ägt en bil (F = förstagångsköpare).

Man fann också att bilköparna ibland köpte ett tilläggspaket med extrautrustning men att det varierade beroende vilken kundkategori de tillhörde. Tabellen nedan sammanfattar resultaten:

Kundkategori (andel)	$Å$ (50 %)	N (40 %)	F (10 %)
T = köpte tilläggspaket	70 %	50 %	30 %
$\bar{T}$ = köpte inte tilläggspaket	30 %	50 %	70 %

a) Givet att en slumpvist vald kund köpte ett tilläggspaket, vad är sannolikheten att kunden är en återvändande kund? Dvs. vad är $P(Å\,|\,T)$? TIPS: Beräkna $P(T)$ först. (5p)

A. 0,580

B. 0,500

C. 0,467

D. 0,603

E. 0,700

Bilhandlaren tittade också på försäljningspriserna och förväntade intäkter. En särskild bilmodell fanns i tre varianter: Premium, Advance och GT. Grundpris och försäljningsandel för respektive variant visas i följande tabell:

Variant	Premium	Advance	GT
Försäljningsandel	50 %	40 %	10 %
Grundpris	190 000	230 000	250 000

b) Ange väntevärde och standardavvikelse för X = grundpriset för en slumpvist vald försäljning. (5p)

A. $\mu_X = 218\,000$ $\sigma_X = 24\,000$

B. $\mu_X = 223\,333$ $\sigma_X = 24\,944$

C. $\mu_X = 230\,000$ $\sigma_X = 24\,944$

D. $\mu_X = 218\,000$ $\sigma_X = 57\,600$

E. $\mu_X = 233\,333$ $\sigma_X = 30\,551$

Uppgift 3

Jens har börjat investera i aktier och har en väldigt enkel strategi för sina aktieköp. En gång per vecka väljer han ut den aktie som han tror ska ge bäst utdelning. Sedan kastar han en vanlig tärning. Om det blir ett jämnt tal så köper han aktien, om det blir udda köper han den inte.

a) Vad är sannolikheten att Jens ska genomföra fler än sju aktieköp under de kommande tolv veckorna? (5p)

A. 0,073

B. 0,806

C. 0,387

D. 0,500

E. 0,194

Jens använder en enkel modell för att bedöma värdet på två olika aktier, AIS och BAC. Han antar att $X =$ värdet på AIS-aktien är normalfördelad med väntevärdet $\mu_X = 25$ kr och variansen $\sigma_X^2 = 81$ och att $Y =$ värdet på BAC-aktien också är normalfördelad men med väntevärdet $\mu_Y = 40$ kr och variansen $\sigma_Y^2 = 121$.

b) Vad är sannolikheten att värdet på AIS-aktien ska överstiga 35 kr? (5p)

A. 0,111

B. 0,134

C. 0,159

D. 0,452

E. 0,867

OBS! Svarsalternativen i a) och b) har avrundats, välj det alternativ som är närmast.

Jens har köpt in sig på $n_X = 20$ AIS aktier och $n_Y = 30$ BAC-aktier. Utöver antagandena ovan antas dessutom att aktievärdena korrelerar negativt med varandra med en korrelationskoefficient $\rho_{XY} = -0{,}40$. Låt W beteckna portföljens totala värde.

c) Vad är väntevärdet μ_W respektive standardavvikelsen σ_W för slumpvariabeln W? (5p)

A. $\mu_W = 1700$ $\sigma_W = 306{,}24$

B. $\mu_W = 1700$ $\sigma_W = 434{,}53$

C. $\mu_W = 65$ $\sigma_W = 16{,}769$

D. $\mu_W = 65$ $\sigma_W = 11{,}082$

E. $\mu_W = 1700$ $\sigma_W = 93780$

Uppgift 4

Som nybliven inköpare på ett större klädesföretag undersöker du kvaliteten i en viss produkt. Leverantören har angett att X = produktens vikt (en slumpvariabel) i genomsnitt är $\mu = 600$ gram med standardavvikelsen $s = 50$ gram. Du planerar att ta ett stickprov av storlek $n = 100$ och sedan kontrollera leverantörens påstående genom att beräkna $\bar{X}$ = stickprovets medelvärde.

a) Givet att leverantörens påstående är sant, vad är sannolikheten att $\bar{X} < 590$ gram ? (5p)

A. 0,003

B. 0,345

C. 0,655

D. 0,977

E. Eftersom vi inte vet fördelningen för X kan sannolikheten inte beräknas, inte ens approximativt.

OBS! Svarsalternativen har avrundats till tre decimaler, välj det som är närmast.

När du dragit ditt stickprov beräknar du stickprovsmedelvärdet till $\bar{x} = 590$ gram. Du utgår ifrån att standardavvikelsen är känd och lika med 50 gram.

b) Vilket av följande alternativ är ett approximativt 95 % konfidensintervall för μ? (5p)

A. (590,2 ; 609,8)

B. (588,6 ; 591,4)

C. (581,8 ; 598,2)

D. (580,2 ; 599,8)

E. (589,9 ; 590,1)

OBS! Konfidensintervallets gränser har avrundats, välj det alternativ som är närmast.

Du tycker att felmarginalen kunde vara mindre än det du fick i b)-uppgiften ovan.

c) Givet att är standardavvikelsen är känd och lika med 50 gram, hur stort stickprov skulle du behöva för att få en felmarginal som är högst lika med 5? (5p)

A. $n = 20$

B. $n = 110$

C. $n = 196$

D. $n = 385$

E. $n = 720$

OBS! Alternativen ovan har avrundats, välj det alternativ som är närmast.

Uppgift 5

Vid ett stort möte med universitetslärare från tre universitetsstäder dracks det en del vin. Två statistiker i sällskapet ville undersöka om de olika universiteten skiljde sig åt i valet mellan rött och vitt vin. Man lyckades samla in data för samtliga 300 deltagare och fördelningen mellan rött och vitt vin för de tre universiteten sammanställdes i en tabell:

Universitet	Rött	Vitt
Lund	50	40
Stockholm	70	30
Uppsala	60	50

a) Vilken typ av test är lämplig i detta fall för att avgöra om det finns skillnader mellan universiteten med avseende på val av vin; och hur är testvariabeln fördelad? (4p)

A. Test för differenser av väntevärden; approximativt normalfördelad (z-test)

B. Test för differenser av parade observationer; t-fördelad med 2 frihetsgrader

C. Oberoendetest (homogenitetstest); χ^2-fördelad med 2 frihetsgrader

D. Oberoendetest (homogenitetstest); χ^2-fördelad med 3 frihetsgrader

E. Anpassningstest (*goodness-of-fit*); χ^2-fördelad med 2 frihetsgrader

b) Bland alternativen nedan anges observerat värde till vänster och kritiskt värde till höger samt slutsats av testet. Givet att du testar på 5 % signifikansnivå, vilket av dessa alternativ anger korrekt observerat värde (värdet i högerledet) på testvariabeln och vilken slutsats är korrekt? (6p)

A. $|-2{,}000| < 3{,}182 =$ kritiskt värde och H_0 förkastas inte, universiteten är lika

B. $5{,}803 < 5{,}991 =$ kritiskt värde och H_0 förkastas inte, universiteten är lika

C. $6{,}271 < 7{,}815 =$ kritiskt värde och H_0 förkastas inte, universiteten är lika

D. $|2{,}449| > 1{,}96 =$ kritiskt värde och H_0 förkastas, universiteten är olika

E. $6{,}271 > 5{,}991 =$ kritiskt värde och H_0 förkastas, universiteten är olika

Fullständig redovisning krävs för följande uppgifter. Använd separata pappersark för uppgift 6 resp. uppgift 7.

Uppgift 6

Ett försäkringsbolag samlade in data från två stickprov av bilförsäkringar; det ena stickprovet bestod av $n_X = 400$ ogifta manliga försäkringstagare, det andra stickprovet bestod av $n_Y = 900$ gifta manliga försäkringstagare. Bland de ogifta hade $X = 76$ inkommit med skadeanmälningar, bland de gifta var motsvarande siffra $Y = 90$. Du får i uppdrag att testa om P_X = andelen ogifta försäkringstagare som har inkommit med skadeanmälningar är större än P_Y = motsvarande andel bland de gifta.

a) Ställ upp lämpliga hypoteser för att testa om andelen P_X är större än P_Y. Ange vilka antaganden som gäller för testet, testvariabel och dess fördelning samt beslutsregel och kritiskt värde. Använd 5 % signifikansnivå. (8p)

b) Genomför testet dvs. beräkna det observerade värdet på testvariabeln och fatta ditt beslut, tolka även slutsatsen i ord. (6p)

c) Förklara kortfattat vad Centrala gränsvärdessatsen (CGS) är och förklara kortfattat hur satsen är relevant för denna uppgift. Max en ½ A4 behövs. TIPS: Hur beräknas $\hat{p}_X$ respektive $\hat{p}_Y$? (6p)

Uppgift 7.

En butikskedja analyserar hur lönsamheten under ett kvartal varierar mellan olika butiker runt om i landet och studerar hur butikernas intäkter samvarierar med olika variabler. I en enkel modell utgår man ifrån att Y_i = intäkten för butik i delvis kan förklaras av X_i = butikens storlek mätt i antal anställda enligt en enkel linjär regressionsmodell:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

I bilagan på följande sida finner du datamaterialet för ett stickprov av storlek $n = 10$ butiker samt några olika beräkningar.

a) Skatta modellens parametrar och tolka de värden du får i ord. (6p)

b) Beräkna R^2 och tolka detta värde, dvs. förklara kortfattat så att din chef förstår. (4p)

c) Beräkna ett 95 % konfidensintervall för lutningskoefficienten. Kan man med hjälp av intervallet dra slutsatsen att lutningskoefficienten är signifikant skild från noll eller inte? Förklara. (6p)

d) Hur ska man tolka ett konfidensintervall för $\mu_{Y|X=x}$ respektive ett prediktionsintervall för $\hat{y}$? Hur förklarar du skillnaden mellan dessa två begrepp för din chef? Räcker med en kort förklaring, max ½ A4 sida. (4p)

Bilaga till Uppgift 7

Data

Obs. i	1	2	3	4	5	6	7	8	9	10	Summa
x_i	3	5	5	7	9	9	11	13	13	15	90
y_i	10	14	12	18	23	18	20	22	21	19	177
x_i^2	9	25	25	49	81	81	121	169	169	225	954
y_i^2	100	196	144	324	529	324	400	484	441	361	3303
$x_i y_i$	30	70	60	126	207	162	220	286	273	285	1719
e_i	-2,45	-0,20	-2,20	2,05	5,30	0,30	0,55	0,80	-0,20	-3,95	

Sammanfattande deskriptiva mått

$\bar{x} = 9$ $\quad s_x^2 = 16$ $\quad \bar{y} = 17{,}7$ $\quad s_y^2 = 18{,}9$ $\quad r_{xy} = 0{,}805076$

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 23/9-2019

Sal: Brunnsvikssalen

Tenta: Statistik för ekonomer

Kurs: Grundläggande statistik för ekonomer

ANONYMKOD: 0035 – CBF

☒ Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X	X	X			3
Lär.ant. 10	10	10	15	10	20	16			

MW

POÄNG	BETYG	Lärarens sign.
91	A	MC

STOCKHOLMS UNIVERSITET HT2019

Statistiska institutionen

SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer 2019-09-23

Anonymkod: 0035-CBF *(skriv tydligt!)*

Markera ditt svar med ett tydligt kryss (X) i rutorna nedan.

OBS! Endast ett kryss per uppgift. Har fler än ett svarsalternativ markerats ges noll poäng.

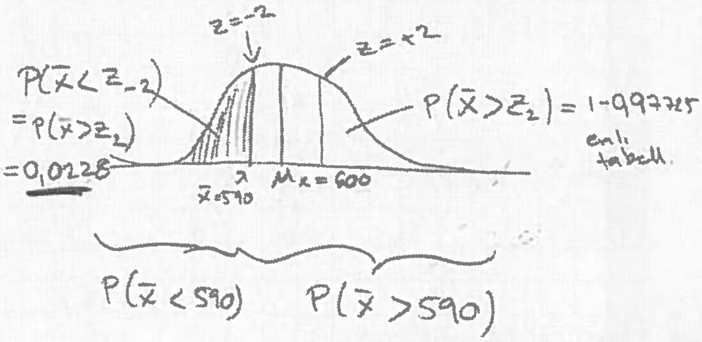
> **OBS! Om du efter att ha kontrollerat dina beräkningar ordentligt kommer fram till att svaret inte finns bland de givna svarsalternativen eller anser att något är fel med uppgiften, skriv ditt svar på vid den aktuella frågan och kommentera på baksidan av detta blad.**

		A	B	C	D	E	
Uppgift 1	a)		X				R
	b)			X			R
Uppgift 2	a)				X		R
	b)						* OBS! Ingen alternativ är rätt, se → nästa sida $\mu_x = 212\,000$:- $\sigma_x = 22\,716$:- R SP
Uppgift 3	a)					X	R
	b)		X				R
	c)					X	–
Uppgift 4	a)						SP * OBS! Jag anser egentligen att → $P(\bar{x} < 590g) = 1 - 0{,}977 = \underline{0{,}023}$ R
	b)				X		R
	c)				X		R
Uppgift 5	a)			X			R
	b)					X	R

2b) $\mu_x = \sum x \cdot P(x) = 190\,000 \cdot 0{,}5 + 230000 \cdot 0{,}4 + 250\,000 \cdot 0{,}1$
$= \underline{212\,000 :-}$

$$\sigma_x^2 = \sum (x-\mu_x)^2 \cdot P(x) = 516\,000\,000, \quad \sigma_x = \sqrt{\sigma_x^2} = 22\,716$$

4a) $z = \dfrac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}} = \dfrac{590-600}{50/\sqrt{100}} = -2{,}0$, $z_{-2,0}$ enl. tabell $= 0{,}97725$



<u>Min slutsats</u>

$P(\bar{x} < 590\,\text{gram}) = 0{,}0228$

$P(\bar{x} > 590\,\text{gram}) = 0{,}97725$

SU, STATISTIK

Skrivsal: BRU **Anonymkod:** 0035-CBF **Blad nr:** 1

Uppg. 6

a) $n_x = 400$, $x = 76$ skadeanmälningar , $\hat{p}_x = 76/400 = 0,19$

$n_y = 900$, $y = 90$ —"— , $\hat{p}_y = 90/900 = 0,10$

Testa: Om $P_x > P_y$

Hypoteser

$H_0: P_x \leq P_y$, $H_1: P_x > P_y$ (Dvs vår nollhypotes, det vi utgår ifrån är att det inte är ngn skillnad, om H_0 förkastas har vi evidens för att $P_x > P_y$)

R 2

Antaganden

iid -> "independent & identically distributed", dvs:

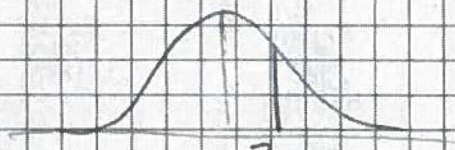
- stickproven tas oberoende av varandra
- sannolikhetsfördelningen/andelen antas vara konstant, (dvs lika) för alla stickprov ok 2-

Testvariabel

fördelning?

$\left.\begin{array}{l} n > 30 \\ \alpha = 5\% \end{array}\right\} \Rightarrow$ Testvariabel $= Z_{0,05} = 1,6449$ R

2-



Beslutsregel

Förkasta H_0 om $Z_{obs} > Z_{krit}$ (dvs 1,6449)

8-

b) Beräkningar

$$Z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}_0 \cdot (1-\hat{p}_0)\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}}$$ a)

$$, \hat{p}_0 = \frac{n_x \cdot \hat{p}_x + n_y \cdot \hat{p}_y}{n_x + n_y} = \frac{400 \cdot 0,19 + 900 \cdot 0,1}{1300} = 0,1277$$

$$= \frac{0,19 - 0,10}{\sqrt{0,1277 \cdot 0,8723 \cdot \left(\frac{1}{400} + \frac{1}{900}\right)}} = \frac{0,09}{\sqrt{0,0004}} = 4,4875$$ R

8

Slutsats

$\underbrace{Z_{obs}}_{4,4875} > \underbrace{Z_{krit}}_{1,6449} \Rightarrow H_0$ förkastast på 5% signifikansnivå.

Dvs det finns evidens på 5% signifikansnivå att P(x) är större än P(y)

6c) Centrala gränsvärdessatsen (CGS) innebär att fördelningen för slumpvariabeln går mot en normalfördelning då $n \to \infty$.
D.v.s. ju fler stickprov vi genomför på en viss population desto mer kommer stickprovsmedelvärdet (i detta fall stickprovsandelen) att närma sig en normalfördelning, även om den underliggande populationen inte är normalfördelad. Våra beräkningar underlättas av att vi kan räkna utifrån en normalfördelning.

I detta exemplet vet vi ju inte om (vi vet att det inte är N) variabeln skadeärenden är normalfördelad bland gifta resp. ogifta män, men tack vare att vi har stora stickprovsunderlag ($n_x = 400$ resp $n_y = 900$) kan vi räkna utifrån att vår variabel är normalfördelad, t.ex $\hat{p}_x = \frac{76}{400}$

/6

20

SU, STATISTIK

Skrivsal: BRU **Anonymkod:** 0035-CBF **Blad nr:** 2

Uppg. 7

a) $y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$, $n = 10$

$$b_1 = \frac{Cov(x,y)}{S_x^2} = \frac{\frac{\sum x_i y_i - n\bar{x}\bar{y}}{n-1}}{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\left.\begin{aligned} \sum x_i y_i &= 1719 \\ n &= 10 \\ \bar{x} &= 90/10 = 9 \\ \bar{y} &= 177/10 = 17{,}7 \\ \sum x_i^2 &= 954 \end{aligned}\right\} = \frac{1719 - 10 \cdot 9 \cdot 17{,}7}{954 - 10 \cdot 9^2} = \frac{126}{144} = \underline{0{,}875}$$

R

$b_0 = \bar{y} - b_1 \cdot \bar{x} = 17{,}7 - 0{,}875 \cdot 9 = \underline{9{,}825}$ R

$\hat{y}_i = 9{,}825 + 0{,}875 \cdot x_i$, b_0 är modellens intercept, d.v.s var linjen skär y-axeln, $\hat{y}_i$ då $x_i = 0$

genomsnitt för y när $x=0$

b_1 är linjens lutningskoefficient. I detta fall är den positiv, d.v.s ju högre x_i desto högre $\hat{y}_i$ → känns rimligt att en större butik har större intäkter.

för varje ökning med 1 i x-led ökar i genomsnitt y med b_1

5

b) $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

$\underline{SSE} = \sum e_i^2 = -2{,}45^2 + -0{,}2^2 + -2{,}2^2 + 2{,}05^2 + 5{,}3^2 + 0{,}3^2 + 0{,}55^2 + 0{,}8^2 + 0{,}2^2 + 3{,}95^2 = \underline{\underline{59{,}85}}$

$\underline{SST} = (n-1) \cdot s_y^2 = 9 \cdot 18{,}9 = 170{,}10$

1− ✓ 1−

$R^2 = 1 - \frac{SSE}{SST} = \frac{59{,}85}{170{,}10} = \underline{0{,}3519} = 35{,}19\%$ 64.8%

Förklaring: R^2 kallas förklaringsgrad och det är ett mått på hur stor andel av den totala variansen (SST) som förklaras av modellen (SSR). I detta fall är förklaringsgraden låg, endast 35,19%, så det ser ut som att en stor del av variationen är slumpmässig.

Det kan också vara så att fler faktorer spelar in, med en multipel regressionsmodell går SSR upp och därmed också förklaringsgraden. (Dock bör då justerad R^2 beräknas för att jämföra modellerna och se till så att de extra variablerna bidrar till bättre prediktion).

3

c) $b_1 \pm t_{n-k-1,\alpha/2} \cdot S_{b_1}$, $t_{n-k-1,\alpha/2} = t_{\underbrace{10-1-1}_{8},\, 0{,}05/2}$ $\alpha/2 = 0.0025$

$= t_{8;\,0{,}025} = 3{,}833$ för stort!

$$S_{b_1} = \frac{S_e^2}{n-1 \cdot S_x^2} = \frac{\Sigma e_i^2 / n-k-1}{(10-1) \cdot 16} = \frac{59{,}85/8}{9 \cdot 16}$$

$= 0{,}0520$ R

$0{,}875 \pm 3{,}833 \sqrt{0{,}0520}$
$\pm 0{,}1991$

KI: $[0{,}6759 ; 1{,}0741]$

Slutsats: Det 95%-iga konfidensintervallet $[0{,}6759; 1{,}0741]$ innehåller ej 0, R dvs på 5% signifikansnivå är lutningskoefficienten signifikant skild från 0, och positiv vilket i sin tur innebär att vi har en positiv lutning på b_1, dvs när x_i ökar så ökar $\hat{y}_i$ enligt modellen. BRA annars! /4

d) Med konfidensintervall för $\mu_{Y|X=x}$ menas att på en viss signifikansnivå så kommer medelvärdet av ett antal ~~x~~ Y att falla inom detta intervall. Tex bland 30 butiker med $X = 5$ anställda så kommer de genomsnittliga y-intäkterna ligga inom detta intervall i 95% av fallen (vid $\alpha = 5\%$) nästan

För prediktionsintervall $\hat{y}_i$ givet $X = x$ menas att en butik med $x = 5$ anställda kan förväntas ha intäkter inom detta intervall i 95% av fallen (vid $\alpha = 5\%$). R BRA!

/4